多変量解析を教育実践家レベルへ

Ⅹ．エクセルを使った　数量化Ⅱ類

2015.06.5　三村紀中

**１．エクセルを活用して学校で使う「数量化Ⅱ類」について**

　数量化Ⅱ類とは、事前に２つのグループに分かれていることが明らかな個体がある時、新しいデータが得られた際に、個体の特性データや回答データを用いて、どちらのグループに入るのかを判別する手法です。判別分析が、量的変数を説明変数とするのに対して、数量化Ⅱ類は質的データ、カテゴリカル・データを説明変数として用います。

　この手法を用いて回帰式の係数を比較し、判別に当たって重視されている要因を分析することによって、影響力の大きな要因を見つけ、新たな知見を得ることもできます。

　分析に当たっては、一般固有値問題を解き、最大固有値に対する固有ベクトルを得ます。得られた固有値は相関比に等しくなります。

　最大固有値に基づく群間の判別が十分でない時は、２番目に大きな固有値（相関比）を用いて固有ベクトルを求めます。同様に、理論上は説明変数の数－１の固有値が得られますが、累積寄与率が８０％を超える程度で計算を終了します。

　大抵は、第１固有値と第２固有値（１軸と２軸）によって０．８以上の累積相関比が得らない場合は、説明変数の構成を見直すことを検討する方が賢明でしょう。

　例題１：テレビ番組を２０個程度示し、それぞれを「良く見る、ときどき見る、あまり見ない、ほとんど見ない」の３段階で回答してもらいます。これを、良く見る＝３点、時々見る＝２点、あまり見ない＝１点、ほとんど見ない＝０点として「簡易視聴率」を集計します。その回答は個人の直観的判断によるものなので、合計点の多寡を細かく分析するほどの妥当性はありません。そこで、視聴率を高・中・低の３区分とし、１群、２群、３群に分けます。

　番組のジャンルを、１：ドラマ（現代劇）　２：ドラマ（時代劇）　３：マンガ　４：クイズ・歌　５：スポーツ　６：旅・情報　７：ニュース　に分けておきます。

　放映時間帯を、１：１９時台以前　２：２０時台　３：２１時台　４：２２時以降に分けておきます。

　これらをクロス表に整理して数量化Ⅱ類で分析すると、番組の内容×放映時間帯の何が視聴率に影響を与えているかを分析・観察することがでます。

例題２：旅館のアンケートの例

「問1. 食事の量について、問2. 大浴場の湯温について、問3. 従業員の接客態度について、問4. 再び宿泊したいと思いますか？」を尋ねたデータがあるとします。

　「再び宿泊したい」と回答したお客さんは、どこが気に入ってくれたのか？　何が重要な要素であったかを調べれば、改善の知見を得ることができると期待されます。

　「再び宿泊したいと答えてくれた人が○○％いた。よかった。」で、終わってしまうようなアンケート処理は、「お客さんの隠れた要望、傾向」というような重要な情報を捨ててしまうようなものなのです。

**２．データの形式、入力の仕方について**

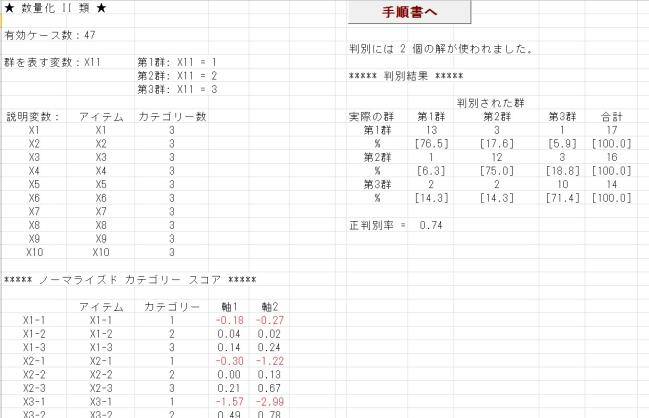
　「属する群の別」を整数値で入力してください。群の数は100個まで対応できます。

　各群に属するケース（回答者）は複数でなければなりません。

　説明変数は複数の質的変数（カテゴリカルデータ）であることが条件です。

　右の図のような画面に、群情報データ、説明変数データ、ケース名などを入力します。データは他のファイルからコピー＆ペーストできます。

　次に「」マクロボタンを押して分析を開始してください。一気に散布図表示まで進みます。

**３．判別関数について**

次に判別関数の一覧を見ます。

　判別式の妥当性は、判別的中率、まはた誤判別率などで評価できますが、的中率は80％以上を目指すべきとされています。

　「重心」は、各群の中心位置を示しています。この例では元の群が３つあるので、第１～第３群の位置が軸１と軸２のベクトル値としてレポートされています。この値は、ケースデータのベクトル値を用いて描いた散布図に重ね合わせることができます。

　「相関比」（群間平方和÷総平方和）は、０から1までの値をとり、値が１に近いほどよく判別され、０に近ければあまり判別されていないことを表しています。相関比はそれぞれの軸の固有値と同じで、レポートされている相関比の総和が累積寄与率になります。

　「偏相関係数」は、外的基準（目的変数＝どの群に属するかの情報）と説明変数の間の相関係数です。レポートされている偏相関係数の総計が大きいアイテム（説明変数）ほど、判別に寄与する度合いが強いと言えます。どの説明変数の影響が大きいかを調べることによって、実践的な知見を得ることができると期待されます。



**４．各ケースの判別の結果について**

　「サンプルスコア」という標題が付けられているのは、各ケース（回答者）の群所属判別の結果と、固有ベクトルの一覧です。各ケースごとに、誤判別されたケースにはマークが付けられています。

　この軸１、軸２の固有ベクトル値を用いて、判別の様子を視覚的に把握、分析することができる散布図を描きます。

**５．判別散布図について**

　多変量解析を実行すると、数学的モデルによる解析計算の結果を得ることができますが、その「意味」や「価値」は、計算内容と調査内容の両方を熟知した複数の人々によって解釈・協議されなければなりません。

　数学的には自動的な言語コメントが得られることはありませんし、誰か一人だけの解釈を鵜呑みにすることもできません。また、多変量解析の専門家、調査者のいずれか一方の意見に偏ることも危険です。

　このブックでは、第１軸×第２軸の固有ベクトルを手がかりにした判別散布図を、群ごとに色分けして表示します。

　軸３以下のレポートが得られた場合は、サンプルスコアのシートから軸３以下の固有ベクトル値を散布図の右側のＸ軸、Ｙ軸の列にコピー＆ペーストすれば、即座にグラフが描き直されます。



**６．おわりに**

　このマクロ・ブックの主計算ルーチンには、群馬大学社会情報学部教授　青木繫伸氏が公開しているＶＢＡマクロを引用しました。厚く感謝申し上げます。

多変量解析の自習ノート：<http://aoki2.si.gunma-u.ac.jp/lecture/mva.html>

多変量解析エクセルVBA：http://aoki2.si.gunma-u.ac.jp/lecture/stats-by-excel-old/vba/index.html

【参考文献等】

・図解雑学　多変量解析　丹慶勝市　2007　ナツメ社

・Excel徹底活用　多変量解析　上田太一郎、他　2003　秀和システム

・意味がわかる多変量解析　石井俊全　2014　ベレ出版

・Interactive 因子分析システム　久本博行　関西大学「社会学部紀要」第34巻第2号,2003

探究的因子分析　久本博行　www2.jpcku.kansai-u.ac.jp/~hisamoto/

・因子分析法　清水和秋　関西大学大学院心理学研究科　www2.jpcku.kansai-u/ac.jp/~shimizu/

・エクセルで統計分析４HADによる因子分析・クラスタ分析　清水裕士　広島大学大学院総合科学研究班

エクセルで統計分析４＿因子分析のやり方.pdf

・重回帰分析、因子分析、主成分分析、クラスター分析、数量化Ⅲ類、判別分析

（青木繁伸　群馬大学社会情報学部）　2003.08.07

統計学自習ノート: http://aoki2.si.gunma-u.ac.jp/lecture/PFA/index.html

エクセルVBA：http://aoki2.si.gunma-u.ac.jp/lecture/stats-by-excel-old/vba/index.html

・因子分析、基本的な考え方、他　心理データ解析Basic　小塩真司　早稲田大学文学学術院

www.f.waseda.jp/oshio.at/edu/data\_b/top.html

・テキストマイニングの手法　島根県農業技術センター　広報誌「ときめき」No.374　2012.1.30

・林の数量化理論って何だ　<http://www.geocities.co.jp/WallStreet/7166/SP/sp_inde.html>

・Albertデータマイニング：クラスター分析の手法（階層クラスター分析）

<http://www.albert2005.co.jp/technology/mining/method3_2.html>